A Novel CAPTCHA Solving Technique Based on Deep Convolutional Neural Networks

Amar U. Khopade
Dept. of Information Technology
Pune Institute of Computer Technology
Pune, India
amarkhopade20@gmail.com

Dr. Emmanuel M.
Dept. of Information Technology
Pune Institute of Computer Technology
Pune, India
emman2001@gmail.com

***Abstract*—CAPTCHA(Completely Automated Public Turing Test to tell Computers Humans Apart) is a computer generated test used to differentiate between human and bots. In this paper we propose an approach that uses deep Convolutional Neural Network(CNN) to integrate localization, segmentation and recognition steps while solving visual text based CAPTCHAs. CNN operates directly on the image pixels. Optical Character Recognition program, In contrast to early work that relied on sophisticated computer vision or machine learning algorithms, we used simple pattern recognition algorithms but exploited fatal design errors that we discovered in each scheme.**

Keywords-CAPTCHA, Convoulutional neural network, OCR, HIPs.

I. Introduction

According to latest study around 60 million CAPTCHAs are solved everyday taking few seconds to decode and type in. Some of handwritten CAPTCHAs are useful in digitizing the handwritten manuscripts and make them available online. With tremendous increase in machine learning and computer automation, there is violation of terms of service and increase in attacks from various sources thus affecting application security[7].

Internet has huge sensitive information which when exposed can be considered as serious criminal activity, so to protect this information many security primitives are developed. CAPTCHA is one of the security primitive widely used in almost all online activities to ensure whether transaction is performed by human or not[8]. Many researcher are engaged in developing techniques to solve CAPTCHAs using various recognition methods to check the possibility of bot attacks.

In computer vision technology object classification is much more challenging than object recognition. For example an image of dog, horse and tree is very clear to human and can be classified easily. However, it turns out that to this day, classification of objects in real-world scenes remains an open and difficult problem[5]. Recognizing known objects, on the other hand, is more tractable, especially when specific shapes undergoing change is easy to predict.

A knowledgeable attacker[1] has shown after breaking popular site eBay audio CAPTCHAs and abusing their registration process that CAPTCHA security cannot be overlooked. It was feasible for attacker using eBay audio CAPTCHA to abuse the registration. Large amount of audio image CAPTCHAS are collected, about 2000 samples of eight CAPTCHAs, and 200 samples for 50 other CAPTCHAs. Paper[1] reserves 200 samples of the first eight CAPTCHAs and all 200 samples of the last 50 CAPTCHAs to evaluate the performance of the solvers. Overall we scraped more than 26,000 CAPTCHAs.

Using hard AI (Artificial Intelligence) problems for security, is an exciting new paradigm which was proposed in[7]. In this paradigm most important creation is CAPTCHA for human and computers by presenting a challenge, i.e., a puzzle, beyond the capability of computers but easy for humans. CAPTCHA is now a standard Internet security technique to protect online email and other services from being abused by bots[2].

IMAGINATION[9] technique has an empirical study of distortion applications which is proposed with varying nature, strength and added human and machine recognisability.

An automated attack to defeat the current state-of-the-art in moving-image object recognition (MIOR) CAPTCHAs is presented[4]. Through extensive evaluation of several thousand real-world CAPTCHAs, this attack can completely undermine the security of the most prominent examples of these, namely those generated by NuCaptcha circa 2012[10].

After examining properties that enable this attack, a series of security countermeasures designed to reduce the success of this attacks, including natural extensions to the scheme under examination, as well as an implementation of a recently proposed idea[6] for which attacks do not appear as readily available are explored. Though text based CAPTCHA is most popular form today, other important CAPTCHA types such as motion based and audio CAPTCHAs are finding their way as new security primitive[12].

Most suitable example of this new type of CAPTCHAs is NuCaptcha [3], which asserts to be “the most secure and usable CAPTCHA,” and serves millions of video CAPTCHAs per day. The general idea explains us that we can decode and recognise motion and test CAPTCHA accordingly. In the case of NuCaptcha, users are shown a video with a series of random codeword moving across a dynamic scene, and solve the CAPTCHA by identifying the characters of the codeword[4].

II. RELATED WORK

The attack in[4] is using feed forward neural network technique for object identification and segmentation tasks. The process in this attack involves several steps such as video stream tracking, foreground extraction, segmentation and classification. There is feedback mechanism primarily used for high and low confidence inferences comparison.

Paper[4] also uses k means clustering to segment the derived trajectories in groups. In classification step neural network is used to classify and trace probability of correct recognition. Both security and usability parameters are very well managed in this approach.

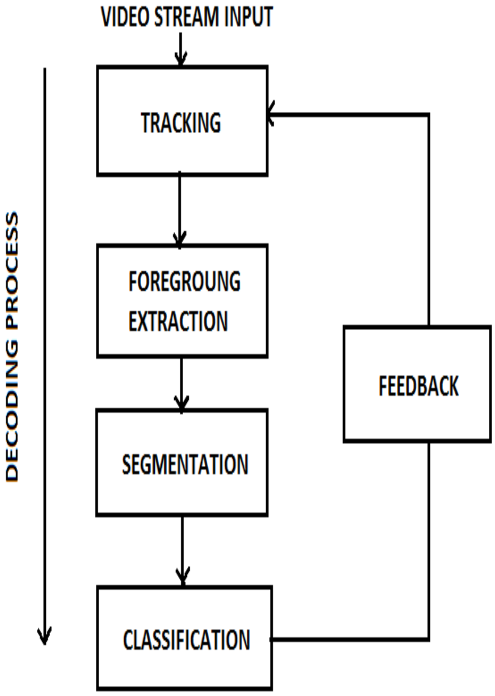

Figure 1. A high level overview of a naive attack[4]

In[14] clickable textual CAPTCHAs are introduced and integrated in a grid. Compare to traditional CAPTCHAs which contained series of distorted characters or strings, the clickable CAPTCHA uses a grid or matrix of specific dimension and allow user to click on mentioned coordinates for authentication or selection of some part of CATCHA space to solve the problem. This paper[14] describes an embodiment of clickable CAPTCHAs based on Google textual CAPTCHAs. 12 Google CAPTCHAs are created and tiled in a 3-by-4 grid. Out of 12, 3 are real English words where as remaining 9 are randomly selected words. To solve the clickable CAPTCHA challenge user must click on the tiles containing only English words.

Paper[15] aims at designing an automatic approach for solving multiple HIPs with minimum intervention using machine learning. Paper explores machine learning method to break top six popular HIPs such as mail blocks, Yahoo v2, Google etc. This method includes custom algorithm designed to locate the characters and machine learning process for recognition[11]. For each HIP both segmentation or recognition steps has around 2500 HIPs trained. Authors encountered a problem during segmentation process. In order to find valid patterns, a recognizer must attempt recognition at many different candidate locations. Segmentation is computationally expensive. Paper concludes that it is efficient to use neural network scheme in the recognition step.

The task in paper[16] is to find the instances of known image objects in a cluttered environment. Objects include words and letters in variable sizes and fonts. Paper[16] two algorithms are proposed algorithm A and algorithm B which are used to break EZ-Gimpy and Gimpy respectively. Two important types of data available during word recognition are lexical information and visual cues.etc.

III. PROPOSED SYSTEM

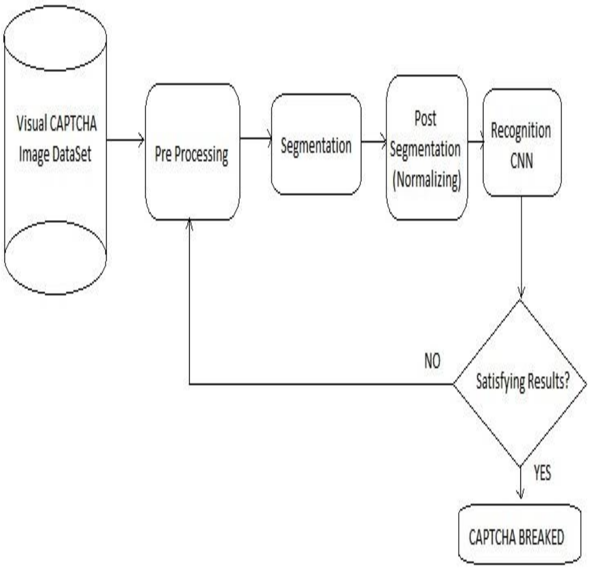

Figure 2. Proposed system diagram.

A. Visual CAPTCHA Image Dataset

JCAPTCHA which stands for Java Completely Automated Public Test to tell Computers and Humans A part is used as application dataset. JCAPTCHA when integrated with web based application gives us variety of CAPTCHAs with multiple characters. JCAPTCHA aims in providing robust and reliable CAPTCHA implementation framework for JAVA, an accessible CAPTCHA implementations and multi-type challenge (text, sound, image).

B. Pre Processing

The Preprocessing will clear the input CAPTCHA image along with this it will convert it into gray scale image, then performs binarization and removes lines or dots if any are present.

C. Segmentation

After preprocessing stage, image characters need to be segmented because it is difficult to recognize characters if they are joined. In JCAPTCHA all the characters are distant so it is comparatively easy to segment them. Checking black pixels to separate characters and recognize becomes easy. The quality of the segmentation depends on the image.

D. Post Segmentation

Segmented image is further processed for noise removal and normalizing the boundary of characters. In this process we are targeting on increasing the font to make the recognition and classification easy. We further talk about the final neural network implementation which will benefit us with accurate recognition and simple model.

E. Convolutional Neural Network(CNN)

Neural networks can be considered as, "Artificial neural networks are generally presented as systems of interconnected neurons which can compute values from inputs, and are capable of machine learning as well as pattern recognition thanks to their adaptive nature."[wiki]. Deep feed forward neural networks have proved use of neural networks in wide range of applications in pattern matching and machine learning.

Convolutional neural networks allow us in understanding how efficiently we can recognize visual image data. There are many classifiers available such as kNN, SVM, softmax which can be used in classification using CNN. CNNs exploit spatially-local correlation by enforcing a local connectivity pattern between neurons of adjacent layers. In other words, the inputs of hidden units in layer m are from a subset of units in layer m-1, units that have spatially contiguous receptive fields.

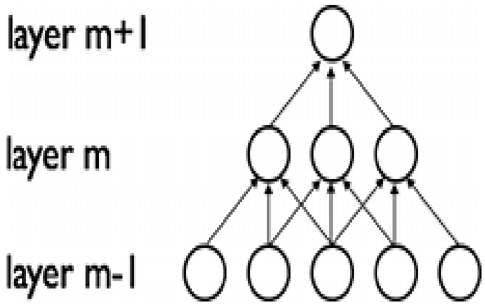

Figure 3. Sparse Connectivity

In addition, in CNNs, each filter h_i is replicated across the entire visual field. These replicated units share the same parameterization (weight vector and bias) and form a *feature map*.

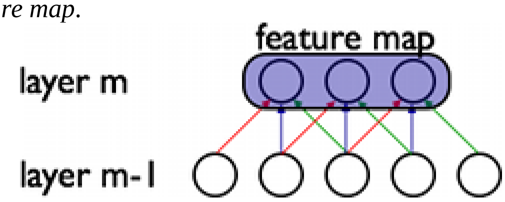

Figure 4. Shared Weights

A feature map is obtained by repeated application of a function across sub-regions of the entire image, in other words, by *convolution* of the input image with a linear filter, adding a bias term and then applying a non-linear function.

IV. CONCLUSION

In this paper a novel approach to solve CAPTCHA using neural network technology is proposed. System consists of block diagram explaining step by step flow of processing and their interconnection. Neural network as compared to other pattern recognition techniques is found efficient and highly accurate in results. Both security and usability primitives can be simultaneously managed using CNN technique.

REFERENCES

[1]. Bursztein and S. Bethard, "DeCAPTCHA: Breaking 75% of eBay Audio CAPTCHAs," Proc. Third USENIX Workshop Offensive Technologies, 2009.

[2]. Bin B. Zhu, Jeff Yan, Guanbo Bao, Maowei Yang, and Ning Xu, "Captcha as Graphical Passwords—A New Security Primitive Based on Hard AI Problems", IEEE Transactions On Information Forensics And Security, Vol. 9, No. 6, June 2014.

[3]. NuCaptcha, "Whitepaper: NuCaptcha & Traditional Captcha," http://nucaptcha.com, 2011.

[4]. Yi Xu, Gerardo Reynaga, Sonia Chiasson, Jan-Michael Frahm, Fabian Monrose, and Paul C. van Oorschot," Security Analysis and Related Usability of Motion-Based CAPTCHAs: Decoding Codewords in Motion", IEEE Transactions On Dependable And Secure Computing, Vol. 11, No. 5, September/October 2014

[5]. S. Ullman, High-Level Vision: Object Recognition and Visual Cognition MIT Press, July 2000.

[6]. N.J. Mitra, H.-K. Chu, T.-Y. Lee, L. Wolf, H. Yeshurun, and D. Cohen-Or, "Emerging Images," ACM Trans. Graphics, vol. 28, no. 5, article 163, 2009.

[7]. L. von Ahn, M. Blum, N. J. Hopper, and J. Langford, "CAPTCHA: Using hard AI problems for security," in *Proc. Eurocrypt*, 2003,pp. 294–311.

[8]. *Yahoo CAPTCHA Hacked*, [Online]. Available: http://it.slashdot.org/article.pl?sid=08/01/30/0037254, retrieved on 01/30/2008, Slashdot

[9]. Ritendra Datta, Jia Li, and James Z. Wang" Exploiting the Human–Machine Gap in Image Recognition for Designing CAPTCHAs", IEEE Transactions On Information Forensics And Security, Vol. 4, No. 3, September 2009.

[10]. Ritendra Datta, Jia Li, and James Z. Wang, "IMAGINATION: A Robust Image-based CAPTCHA Generation System", *MM'05,* November 6–11, 2005, Singapore.

[11]. Miller, "WordNet: A Lexical Database for English,"Comm. of the ACM, 38(11):39-41, 1995.

[12]. N. Friedman and S.J. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach," CoRR (technical report) vol. abs/1302.1539, http://arxiv.org/abs/1302.1539, 2013.

[13]. Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet," Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks",Street View and reCAPTCHA Teams, Google Inc 14 Apr 2014.

[14].]. R. Chow, P. Golle, M. Jakobsson, L. Wang and X. Wang, "Making CAPTCHAs Clickable" In proc. of *HotMobile 2008*.

[15]. [2]. K. Chellapilla and P. Simard. Using machine learning to break visual human interaction proofs (HIPs). *Advances in Neural Information Processing Systems 17*, 2004.

[16]. [3]. G. Mori and J. Malik. Breaking a visual captcha. *Computer Vision and Pattern Recognition*, 2003.